HHS Public Access

Author manuscript

J Registry Manag. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

J Registry Manag. 2016 ; 43(1): 10–15.

Validating Procedures used to Identify Duplicate Reports in Haiti's National HIV/AIDS Case Surveillance System

Chris Delcher, PhD[a], **Nancy Puttkammer, PhD**[b], **Réginald Arnoux, MD**[c], **Kesner Francois, MD**[d], **Mark Griswold, MSc**[e], **Irum Zaidi, MPH**[f], **Yves Anthony Patrice Joseph, MD, MPH**[f], and **Barbara J. Marston, MD**[f]

[a]University of Florida, Gainesville, Florida

[b]International Training and Education Center for Health, Seattle, Washington

[c]International Training and Education Center for Health, Haiti (I-TECH Haiti), Port-au-Prince, Haiti

[d]Ministère de la Santé Publique et de la Population, Port-au-Prince, Haiti

[e]National Alliance of State and Territorial AIDS Directors, Washington, DC

[f]Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Objectives—Valid deduplication of human immunodeficiency virus (HIV) case reports is critical to the utility of these data to inform HIV programs. The Haitian Ministry of Health (MSPP) and partners operate a case-based, national HIV/AIDS surveillance system (HASS), using deterministic and probabilistic procedures to identify duplicate records. These procedures are described and validated based on expert classifications.

Methods—Two samples of HASS records identified as duplicates were selected: 100 pairs from deterministic and 100 pairs from probabilistic matching procedures (total: 200 pairs, 400 case reports). Clinical data from the national electronic medical record (iSanté) were reviewed and consensus gold-standard determinations on the status of duplications were made. False positive rates (FPR) were estimated by reviewing these records, while false negative rates were calculated (FNR) by using LinkPlus™ probabilistic linkage software. The effect of deduplication on total HIV case counts was demonstrated.

Results—Review of deterministic matches yielded 99 true positives and 1 false positive (FPR, 1 per 100; 95% CI, 0.71–5.4). Review of probabilistic matches yielded a FPR of 6 per 100 (95% CI, 2.7–12.4). LinkPlus identified 1,491 probable matches among 68,393 records, representing a FNR of 2 per 100 (95% CI, 0.55–7.0). After adjustment, the estimated unique count of reported HIV patients in HASS was 211,885 (95% CI, 207,293–213,232) as of December 2013.

Conclusions—Based on application of the established procedures, HASS conforms to the duplication performance standard recommended by the Centers for Disease Control and Prevention for HIV surveillance.

Address correspondence to Chris Delcher, PhD, University of Florida, 1329 SW 16th St, Rm 5130, PO Box 100177, Gainesville, FL 32610-0177. Telephone: (352) 265-0111 ext. 86238. Fax: (352) 265-8047. cdelcher@ufl.edu.

Keywords

deduplication; Haiti; HIV surveillance

Introduction

The Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that 140,000 people are living with human immunodeficiency virus (HIV) in Haiti as of 2013.[1] The Haitian Ministry of Health (MSPP) implemented a national HIV/AIDS case-based surveillance system (HASS) in 2008, which has produced estimates of new HIV diagnoses and other indicators for monitoring the status of the epidemic.[2,3] In 2011 alone, HASS contained approximately 23,000 newly reported cases of HIV.[4] HIV surveillance, from population counts to continuity of care, relies on accurate, unique identification of patients from case reports to reduce the possibility of duplication and/or improper data linkage.

In countries lacking unique identifiers, duplication rates can be high. Before Brazil implemented a national unique identifier in 2009, it was estimated that only 100 million people were represented in the 140 million records in one of the primary health information systems.[5] Despite extensive discussion about implementing unique identification numbers in Haiti, progress has been delayed. Although many adults are assigned unique identification numbers for voting purposes, 53% of the population is ineligible to vote, and 19% of eligible voters are unregistered.[6] Other proposed approaches have included biometric identifiers (eg, finger scans), health passports, and portable electronic medical records.[5,7]

During HASS's development and piloting phases, estimates showed the probability of duplication was high given the lack of a unique national identifier, clinic-to-clinic patient mobility in Haiti[8] combined with limited inter-clinic communication, and stigma or incentives that would encourage the provision of false information. The MSPP partially addressed these concerns by supporting name-based HIV reporting to increase the likelihood of accurate epidemiologic counts, and applying deterministic and probabilistic patient-matching algorithms using names and other demographic variables.

HASS receives HIV case reports from all facility-based venues where HIV testing and counseling (HTC) services are provided, including HTC (formerly referred to as voluntary counseling and testing) sites, preventing mother-to-child transmission sites, and *tuberculosis*/HIV clinics, via both the MSPP's Monitoring, Evaluation and Surveillance Interface (MESI) reporting system and from 3 centralized clinical systems. These systems are the iSanté electronic medical record (EMR) system, the Haitian Group for the Study of Kaposi's Sarcoma and Opportunistic Infections (GHESKIO), and Partners in Health Zanmi Lasante. The iSanté EMR system is the MSPP's primary EMR for the national HIV care and treatment program. iSanté was deployed in Haiti in 2005 and is currently in use in 98 urban and rural facilities located in all 10 administrative departments in Haiti.[9,10] Together, the 3 systems support clinical services for the majority of people living with HIV in Haiti. At the end of 2012, MESI, iSanté, GHESKIO, and Partners in Health Zanmi Lasante had submitted approximately 31%, 37%, 22%, and 10%, respectively, of the case reports received by HASS.

HASS is operating nationally and there is increasing interest in using HIV surveillance data in Haiti to monitor program/health facility performance and impact. To understand the accuracy of the case counts in HASS, the deterministic and probabilistic approaches to record deduplication were reviewed and validated.

As data are centralized from these providers, data fields are processed prior to deduplication in HASS (eg, special characters like accent marks removed, name abbreviations such as “JN” converted to “JEAN”). During preprocessing, a pseudo-unique, 7-digit HIV reporting code (which is also manually entered from the paper-based forms) is automatically generated for each record. The code consists of the following 7 characters: first and surname initials, birth month and year, and the first initial of the mother’s first name (eg, XY0175Z).

Identifying duplicate records within HASS begins with a series of deterministic matches using first and surname, birth month and year, sex and mother’s first name, the first 4 letters of the patient’s first name, the reporting clinic, and the birthplace. Records matching exactly in all of these fields are automatically assigned to the same patient; missing values are not permitted to match.

The next step is human adjudication. This process is described as “probabilistic” due to the uncertainty associated with the human pattern recognition required to determine if the patient is the same person.[11] The process starts by displaying records with the same pseudo-unique HIV case reporting code created in preprocessing, or same first and last name on the secure HASS website for visual inspection by epidemiologic staff in Haiti. Staff then decide if records represent the same or different persons based on variables such as the patient’s first and surname, mother’s maiden name, sex, birthdate, commune/department of residence, commune of birth, marital status, occupation, date of HIV diagnosis, reporting clinic, reporting system, and report date.

Methods

Validation of Deduplication Procedures

The national EMR system (iSanté) was selected for the validation exercise because of its programmatic and geographic representativeness. Due to limited resources and for ease of calculation, we selected 100 matched pairs of records generated from the deterministic routine and 100 identified from the probabilistic process to review and selected matches with exactly 2 possible records in HASS.

Analyses

Figure 1 provides an overview of the record review process. To evaluate the matches identified by HASS matching procedures, we used a 2-stage expert review process: (1) a central-level, administrative review and (2) a local-level, physician review for final adjudication. The administrative reviewer has worked with the iSanté EMR system since 2005. The physician reviewer is an internal medicine specialist with 7 years’ experience providing HIV care and 5 years as an HIV clinical trainer and mentor in Haiti.

The evaluated data fields and stepwise process used for administrative and physician review are shown in Table 1. When administrative review (Step 1) could not confirm a match using the fields shown, we asked for local physician review (Step 2). The physician had access to the complete longitudinal clinical record for each patient from the iSanté EMR.

Subjective interpretation was allowed. If differences between records were judged plausible given factors such as clinical measurement error, data entry errors, etc., they were accepted as matches. Examples are shown in Table 1.

We calculated false positive rates (FPRs) for the deterministic and probabilistic matched pairs using the final expert determination as the gold standard; 95% confidence intervals were calculated using the score interval method.[12] For the calculations, matches unclassifiable by expert review were divided equally as matches and nonmatches.

To identify additional matches not identified by the HASS matching procedures, we used LinkPlus™, a probabilistic record-linkage software developed by the Centers for Disease Control and Prevention (CDC)'s Division of Cancer Prevention and Control.[13] Designed to help cancer registries detect duplicate case reports, the expected inputs are data elements commonly found in disease registries (eg, first and surname, gender, race/ethnicity, US Social Security number), but the CDC indicates that LinkPlus can be used with "any type of data."[13] A recent study identified possible matches between patients attending different antenatal care clinics in Senegal using LinkPlus.[14]

LinkPlus (default settings) was used to identify possible matches among records that had not been detected as possible matches upon submission to HASS or been adjudicated as nonmatches by HASS surveillance staff. We matched on first, last, and middle name; birthdate; and sex. First and surname were used as blocking variables, increasing linkage efficiency in large data set, as suggested by the LinkPlus manual.[15] The name-based matching used the Jaro–Winkler metric, comparing agreement between 2 strings accounting for random insertion, deletions, and transpositions.[16] Birthdate matching accounted for the absence of in the month, day, and year elements. Sex had to match exactly.

Using a subset of iSanté records, we calculated a possible false negative rate (FNR) as the maximum number of pairs of duplicate records identified by LinkPlus divided by the total number of records considered to be unique following application of HASS deduplication procedures; 95% confidence intervals were calculated using the score interval method.[12] Resource limitations prevented a detailed review of all matches identified by LinkPlus.

Applying Deduplication Rates to HASS

Our goal was to evaluate the effect (with confidence intervals) of deduplication on the total case counts in HASS. We compared the unadjusted number of cases reported to HASS (since December 2013) with adjusted counts using 4 different deduplication approaches: (1) deterministic matching using pseudo-unique HIV reporting codes manually entered from case reporting forms, (2) deterministic matching after basic data quality control to correct/generate missing reporting codes (eg, if the reporting code was missing birth month, we used

birth month from the birth-date field), (3) deterministic and probabilistic matching results from HASS, and (4) the results from the latter adjusted for the estimated FPR and FNR.

Results

False Positive Rate from Deterministic Matches

Based on administrative review, 94 pairs identified by the deterministic procedures were considered true positive matches, 1 was a possible false positive match, and the validities of 5 were undetermined. After final adjudication, the 5 undetermined matches were identified as positive matches; thus, the totals were 99 true positives and 1 false positive, and the estimated false positive rate (FPR) was 1 case per 100 (95% CI, 0.71–5.4). The false positive result arose because fields in that patient's EMR used for matching in HASS were updated after the case had been reported to HASS. After reviewing the updated record, it was clear the 2 patients were different, but currently no process for receiving retrospective updates in HASS exists.

False Positive Rate from Matches Made by Human Adjudication

After administrative review, 84 pairs identified by the probabilistic procedures were considered true positive matches, 2 were possible false positive matches, and the validities of 14 were undetermined. After final adjudication, the counts were 91 true positive matches, 3 false positives, and 6 undetermined matches. After applying the assumption that half undetermined matches were true matches, the estimated FPR was 6 cases per 100 (95% CI, 2.7–12.4).

False Negative Rate from LinkPlus

At the time of the validation exercise, HASS contained 68,393 records submitted from the iSanté EMR that did not match via the deterministic component or had been adjudicated during review as nonmatches. LinkPlus identified 1,491 probable matches in this group (FNR, 2.2 per 100 records [95% CI, 0.55–7.0]). As viewed on the LinkPlus user interface, the majority of these records appeared to be different patients that should not be matched.

Applying Estimates to HASS

By December 2013, HASS contained 302,718 HIV/AIDS case notification records from 4 reporting systems. Using current deterministic and probabilistic matches from HASS, the estimated unique patient count was 213,318. Of the 213,318 unique patients, there were 153,065 (72%) patients with 1 case notification record submitted; 41,059 (19%) with 2 records submitted; 12,129 (6%) with 3 records submitted; 4,177 (2%) with 4 records submitted; and 2,888 (1%) with 5 or more records submitted.

Among the 60,523 patients with 2 or more notification records, 39,741 and 20,512 patients were identified based on the deterministic and probabilistic matching, respectively. Application of the FPR from the deterministic validation component (0.01) to 39,741 patients suggests that 397 (95% CI, 278–2,146) were erroneously matched. Likewise, application of the FPR from the probabilistic validation component (0.06) suggests that 1,231 (95% CI, 551–2,543) were possibly erroneously matched. Based on application of the

estimated FNR from the LinkPlus component (0.02) to the 153,065 patients presumed to be unique records, 3,061 (95% CI, 918–10,715) duplicate case reports may have gone undetected.

The case count would have been 167,954 had deduplication been based solely on the pseudo-unique reporting code. When data quality of the reporting code improved, the unique case count increased to 213,117, partially attributable to fewer missing reporting codes (and thus fewer instances of records assumed to be unique based on poor-quality codes).

Combining these adjustments, we obtained an estimate of 211,885 (95% CI, 207,293–213,232) total unique case counts since December 2013. Table 2 shows the impact of applying different adjustments based on our validation review components, and Figure 2 shows the impact on estimated total unique HIV case counts under each scenario.

Discussion

Case-based surveillance is one component of a country's HIV surveillance activities that can provide critical information about HIV epidemics in many regions of the world.[17,18] These ongoing data collection systems should be evaluated periodically to ensure they meet design objectives, including detection of duplicate case reporting. Reliable case counts may depend on the application of reasonable/validated approaches to identifying duplicate reports, particularly in settings where individuals are not assigned national unique identification numbers. Several deduplication procedures used by the national HIV/AIDS Surveillance System in Haiti were evaluated and validated. After deduplication, the system conforms to recommended CDC duplication performance standards.[19] Without deduplication, the number of HIV case reports purported to represent a single person in the HASS system is markedly overestimated.

The case count attained from identifying duplicates based on the pseudo-unique reporting code is similar to the case count following deterministic and probabilistic matching. However, now we have evidence that the deterministic/probabilistic matching is considerably more accurate at the patient level than the matching based only on the pseudo-unique code at the population and case level.

When deciding how best to deduplicate surveillance data, factors to consider include the data flow point where deduplication occurs, software environment, case volume, data quality, and staffing resources. Deterministic matching can be accomplished in many off-the-shelf or specialized database applications and packages, including Microsoft Access, SQL Server, and EpiInfo.[20] However, extracting data from disparate data systems, preprocessing, and customizing deterministic algorithms require staff with computer programming skills. Human adjudication may identify matches missed by deterministic algorithms but may be inconsistent and burdensome. Thus, human review may be preferable when caseloads are manageable.

In countries like Haiti, where HIV case volumes are high and staffing resources are low, identifying ways to reduce the burden of human adjudication is important. For example, data fields were identified that may improve the ability to automatically discriminate between

true and false matches, thereby reducing the pool of possible records requiring adjudication. These data fields include core variables already captured for HIV surveillance in Haiti (eg, patient full address, phone number, pediatric vs adult patient) and clinical variables in the iSanté EMR that could be incorporated in the future (eg, emergency contact name, most recent height, date and reason for discontinuation of treatment, evidence of transfer in or out). Other less resource-intensive approaches could include simple modifications of the manual review screen to increase the speed of pattern recognition (eg, color-coding mismatching fields between records).

Preliminary review of possible duplicates detected by LinkPlus indicated many were unlikely to be true duplicates. It is important to note that MSPP does not intend to adjust estimated surveillance totals based on the duplications identified by LinkPlus. The usefulness of LinkPlus will vary by country as the software was designed to work best with specific inputs (eg, names written in English, US Social Security numbers).

This review has important implications. First, we have shown that HASS provides an accurate, acceptable approach to patient record matching without a national identifier in Haiti. This finding should improve the confidence in the internal validity of future surveillance data. However, evaluation of other components of the surveillance system needs work. To examine external validity, one must make comparisons between findings from HASS and estimates from other sources. Second, we have identified a replicable set of algorithms and processes from health information systems in Haiti collecting the same set of identifiers. The methodology is being shared and evaluated within Haiti and other countries interested in case-based surveillance. Third, accurate patient matching allows us to understand and improve other aspects of longitudinal patient-level outcomes analyses such as transfer-adjusted analyses of patient retention.[21] For example, Delcher et al (2012) reported that between 2006–2012, approximately a quarter of female patients originally diagnosed and reported from a GHESKIO-supported health facility subsequently visited another clinic outside the GHESKIO network where an additional case report was generated.[2] In a different study, GHESKIO researchers incorporated a HASS look-up step to understand patient transfer patterns. They found that, after accounting for transfers to facilities outside of the GHESKIO network, estimated 24-month, lost-to-follow up rates changed from 52% to 43%.[22] MSSP is establishing procedures for sharing deidentified information for patients in iSanté and other systems. Benefits of increased data sharing include consolidation of records for improved retention. Fourth, accurately matching patient records allows HASS to provide a clearer picture of HIV patient care from point of diagnosis through treatment by using the best information available across information systems. For example, initial data from counseling and testing case reports can provide robust risk-factor information, while EMR-based case data provides longitudinal data such as CD4 cell counts, antiretroviral treatment regimens, and other clinical variables.

This assessment has several limitations. We selected our sample for convenience and only reviewed pairs with 2 possible matching records from the iSanté system. Thus, our findings may not be generalizable to the other EMR systems reporting to HASS. Further evaluations are needed to understand the validity of the matching algorithms for the full combined data set or other situations. Second, adjudication by local Haitian staff was used as the gold

standard for this evaluation. This adjudication is subjective, and we have not validated the accuracy or reliability of their decisions. We are in the process of formally documenting the decision-making logic. Third, we used LinkPlus, a US system not developed to identify duplicate entries among previously deduplicated records in Haiti.

Our findings may not apply to systems with variable levels of data quality or different cultural practices (eg where similar names are more or less common or people are more or less likely to accurately report information such as birthdate). Additionally, the record-consolidation process used for surveillance and program evaluation purposes should be used with caution for patient management at the clinic-level.

Conclusion

In conclusion, the matching approach yielded an acceptable error rate for national-level HIV/AIDS surveillance purposes. These findings are being used to improve the accuracy of case reporting in Haiti. We recommend that countries develop strategies to prevent case duplication, especially when national identifiers are unavailable; iteratively test the strategy in coordination with local experts prior to scale-up; and periodically validate and modify the matching process.

References

1. Haiti: HIV and AIDS Estimates. UNAIDS; 2013. website. http://www.unaids.org/en/regionscountries/countries/haiti/
2. Delcher C, Meredith G, Griswold M, et al. Lost to follow-up but perhaps not lost in the health system. J Acquir Immune Defic Syndr. 2012; 61(5):e75–e77. [PubMed: 23183154]
3. Meredith G, Roussel B, Delcher C, Duval N, Louissaint E. Building an HIV/AIDS Case Surveillance System. Paper presented at: Caribbean HIV Conference. 2011
4. Diagnoses of HIV Infection and AIDS in the United States and Dependent Areas. HIV Surveillance Report. Vol. 23. Centers for Disease Control and Prevention; 2011. website. http://www.cdc.gov/hiv/library/reports/surveillance/2011/surveillance_Report_vol_23.html
5. Joint United Nations Pogramme on HIV/AIDS (UNAIDS). Developing and Using Individual Identifiers for the Provision of Health Services including HIV: Proceedings from a Workshop, 24–26 February 2009, Montreux, Switzerland. http://www.unaids.org/en/media/unaids/contentassets/documents/dataanalysis/20110520_Unique_Identifiers_Meeting_Report_Montreux.pdf
6. Voter turnout data for Haiti. International Institute for Democracy and Electoral Assistance; 2011. website. http://www.idea.int/vt/countryview.cfm?CountryCode=HT
7. Buluswar, S.; Friedman, Z.; Mehta, P.; Mitra, S.; Sathre, R. 50 Breakthroughs: Critical Scientific and Technological Advances Needed for Sustainable Global Development. Berkeley, CA: Institute for Globally Transformative Technologies Lawrence Berkeley National Lab; 2014. https://ligtt.org/sites/all/files/page/50BTs-Consolidated.pdf
8. Duval, N.; Meredith, G.; Delcher, C.; Roussel, B. Uniting and unifying existing HIV/AIDS case data to profile patient duplication, mobility and access to treatment and care services. Paper presented at: International AIDS Conference; July 2012; Washington DC.
9. Matheson AI, Baseman JG, Wagner SH, et al. Implementation and expansion of an electronic medical record for HIV care and treatment in Haiti: an assessment of system use and the impact of large-scale disruptions. Int J Med Inform. 2012; 81(4):244–256. [PubMed: 22361158]
10. Lober WB, Quiles C, Wagner S, et al. Three years experience with the implementation of a networked electronic medical record in Haiti. AMIA Annu Symp Proc. 2008:434–438. [PubMed: 18999283]

11. Clark DE. Practical introduction to record linkage for injury research. Inj Prev. 2004; 10(3):186–191. [PubMed: 15178677]
12. Washington State Department of Health. Guidelines for Using Confidence Intervals for Public Health Assessment. Olympia, WA: Washington State Department of Health; 2012. http://www.doh.wa.gov/Portals/1/Documents/5500/ConfIntGuide.pdf
13. Registry Plus™ Link Plus. Centers for Disease Control and Prevention; website. http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm
14. Dellicour S, Brasseur P, Thorn P, et al. Probabilistic record linkage for monitoring the safety of artemisinin-based combination therapy in the first trimester of pregnancy in Senegal. Drug Saf. 2013; 36(7):505–513. [PubMed: 23673815]
15. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidemiol. 2002; 31(6):1246–1252. [PubMed: 12540730]
16. Winkler, WE. Overview of Record Linkage and Current Research Directions. Washington DC: Statistical Research Division, US Census Bureau; 2006.
17. Evaluating a national surveillance system: WHO/UNAIDS Working Group on Global HIV/AIDS and STI Surveillance. World Health Organization; website. http://www.who.int/hiv/pub/surveillance/2013package/module7/en/
18. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep. 2004; 53(RR-5):1–11.
19. Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. Centers for Disease Control and Prevention. MMWR Recomm Rep. 1999; 48(RR-13):1–27. 29–31.
20. Newman LM, Samuel MC, Stenger MR, et al. Practical considerations for matching STD and HIV surveillance data with data from other sources. Public Health Rep. 2009; 124(suppl 2):7–17. [PubMed: 27382649]
21. Jonas, A.; Maher, AD.; Shiraishi, et al. National ART Monitoring System Enables Robust Transfer-Adjusted Retention Analyses in Namibia. Paper presented at: CROI Conference; March 2014; Boston, MA.
22. Noel E, Esperance M, McLaughlin M, et al. Attrition from HIV testing to antiretroviral therapy initiation among patients newly diagnosed with HIV in Haiti. J Acquir Immune Defic Syndr. 2013; 62(3):e61–e69. [PubMed: 23254154]

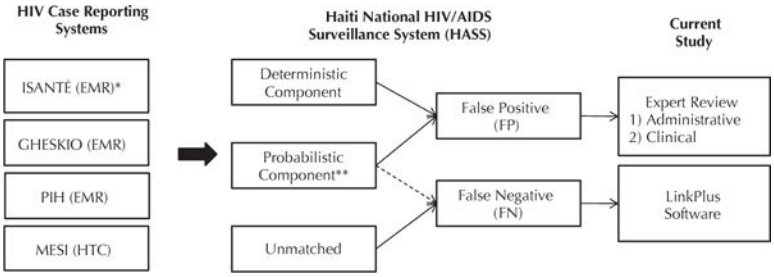

Figure 1. Overview of HIV Case Reporting and the Current Validation Study of Patient Identification in Haiti's National HIV/AIDS Surveillance System

EMR, electronic medical record; HIV, human immunodeficiency virus; HTC, HIV testing and counseling; GHESKIO, Haitian group for the study of Kaposi's sarcoma and opportunistic infections; PIH, Partners in Health Zanmi Lasante. *Isanté and the Haitian Ministry of Health (MSPP)'s Monitoring, Evaluation and Surveillance Interface (MESI) operate in parallel in most MSPP clinics. **Dashed line indicates that records adjudicated as nonmatches by local Haitian staff were not validated in this study. Rather, these records are in the pool analyzed using LinkPlus.

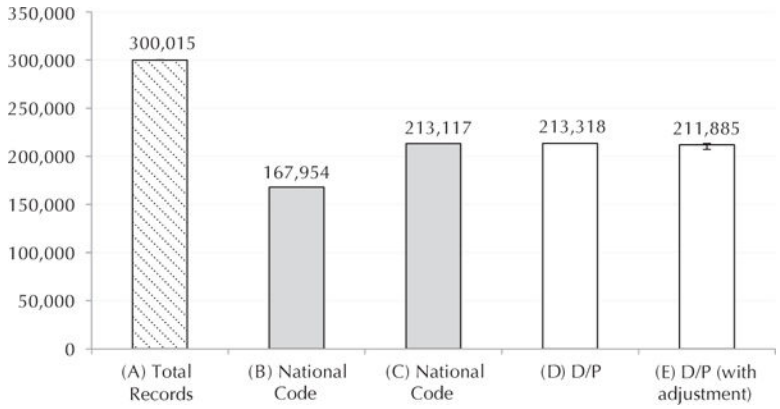

Figure 2.

Comparison of Total Records Submitted to Haiti's National HIV/AIDS Surveillance System (A) Versus Counts Using (B) the National Code "As Is", (C) the National Code Cleaned, (D) Deterministic/Probabilistic (D/P) Methods Currently Used, and (E) D/P Methods after this Validation Review

Table 1

iSanté Data Elements Used for Administrative and Clinical Review in this Study

Expert Review	*Data Elements Reviewed (X = Shared Element with HASS)*	*Data Presenting Ambiguity (Uncertain Cases)*	*Matching Logic*	
			Match	**No Match**
Administrative	• First, last name (X) • Mother's first name (X) • National ID (X) • Clinic (X) • Gender (X) • Date of birth (X) • Total number of clinic visits (X) • First HIV test date (X) • Distinct height measures (with date) • Most recent weight measure (with date) • Earliest clinic visit date and encounter type • Most recent clinic visit date and encounter type • Most recent CD4 (with date) • Earliest ART regimen (with date) • Most recent ART regimen (with date) • Most recent gravida/para/living children (GPEV, with date) • Discontinuation (with reason and date)	• Different gender/Mother's first name different • Date of birth with divergence in some elements of dd/mm/yy • Inconsistent information on visit history and date of death • Heights inconsistent • GPEV divergent	• Date of birth with little or no divergence in dd/mm/yy • Similar mother's first name • Similar height for adult or height trajectory for child • Visit histories compatible • Similar GPEV with plausible differences over time	• Mother's first name different • Date of birth with divergence in all elements of dd/mm/yy • Height widely different during similar time period • Weight widely different during similar time period • GPEV widely different
Physician (same as administrative and:)	• Place of birth (free text) • Address (free text) • Emergency contact name (free text) • Spouse/partner name (free text) • Referral and transfer data (coded and free text) • Reason for discontinuation (free text) • Weight trajectory (coded data) • CD4 trajectory (coded data) • Clinical encounter data (coded and free text)	• Place of birth or current place of residence distinct, but phonetically similar • Weight trajectory similar, but height different • Limited encounter and clinical information to make judgment	• Address or place of birth • Spouse/partner name • Transfer data consistent • Dates of death consistent • Consistent prior treatment history when transferring in • Consistent CD4 trajectory • Inconsistent visit history explained by data entry errors in visit dates/"True" gender	• Adult vs pediatric patient • Place of birth is different • Name of emergency contact different

Expert Review	*Data Elements Reviewed (X = Shared Element with HASS)*	*Data Presenting Ambiguity (Uncertain Cases)*	*Matching Logic*	
			Match	**No Match**
			verifiable in the record by provider notes	

ART, antiretroviral therapy; CD4, cluster of differentiation 4; GPEV, gravidity, parity, enfants vivants; HASS, HIV/AIDS surveillance system; HIV, human immunodeficiency virus; ID, identification.

Table 2

Estimated Count of Unique Patients Reported to the National HIV/AIDS surveillance system in Haiti, as of December 2013

		Patient Counts		
Match Type (Rate Type)	**Error per 100 (95% CI)**	**Current Estimate (No.)**	**Adjustment (95% CI)**	**Adjusted Estimate (95% CI)**
Deterministic (FP)	1 (0.7–5.4)	39,741	+397 (278–2,146)	40,138 (40,019–41,887)
Probabilistic (FP)	6 (2.7–12.4)	20,512	+1,231 (554–2,543)	21,743 (21,066–23,055)
Unmatched (FN)	2 (0.6–7.0)	153,065	−3,061 (−10,715 to −918)	150,004 (142,350–152,147)
Total		213,318	−1,433 (−6,025 to −86)	211,885 (207,293–213,232)

FP, false positive; FN, false negative.